In situ structural restraints from cross-linking mass spectrometry in human Mitochondria

In Situ Structural Restraints from Crosslinking Mass Spectrometry in Human Mitochondria

Petra S. J. Ryl[1], Michael Bohlke-Schneider[1,#], Swantje Lenz[1], Lutz Fischer[1,2], Lisa Budzinski[1], Marchel Stuiver[1], Marta M. L. Mendes[1], Ludwig Sinn[1], Francis J. O'Reilly[1] and Juri Rappsilber*[1,2]

[1] Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

[2] Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, United Kingdom

[#] M. B-S. performed this work before joining Amazon

*Corresponding Authors E-Mail: juri.rappsilber@tu-berlin.de

Abstract

The field of structural biology is increasingly focusing on studying proteins *in situ*, i.e. in their greater biological context. Crosslinking mass spectrometry is contributing to this effort, typically through the use of MS-cleavable crosslinkers. Here, we apply the popular non-cleavable crosslinker disuccinimidyl suberate (DSS) to human mitochondria and identify 5,518 distance restraints between protein residues. Each distance restraint on proteins or their interactions provides structural information within mitochondria. Comparing these restraints PDB deposited structures and comparative models detects novel protein conformations. Our data suggest amongst others substrates and protein flexibility of mitochondrial heat shock proteins. Through this study we bring forward two central points for the move of crosslinking mass spectrometry towards large-scale *in situ* structural biology: clustered better than error-rich individual conflicts of crosslink data with other structural data reveals *in situ* protein conformation states and non-cleavable crosslinkers are compatible with proteome-wide studies. This opens the field for *in situ* photo-crosslinking or other functionalised crosslinkers which were excluded so far from large-scale studies.

Key words: crosslinking mass spectrometry, *in situ* large-scale structural biology, non-cleavable DSS crosslinker, human mitochondria, comparative modelling

Introduction

Mitochondria are complex organelles that fulfil a wide set of essential cellular functions including energy metabolism in all eukaryotic cells [1]. Damaged and dysfunctional mitochondria are implicated in several metabolic, cardiovascular and neurological disorders, and also cancer [1–5]. To fully understand the molecular basis of mitochondrial physiology and its role in disease, it is essential to identify all the relevant components and to reveal their structure and interactions. Human mitochondria have 1,157 proteins currently annotated in MitoCarta 2.0 [6], for fewer than 300 of these we found structures deposited in the PDB, often only covering fragments of the proteins. Commonly used structural biology techniques usually require purification of proteins, which may compromise their structure, solubility or stability [7–9]. Ideally structure elucidation is done in the protein's native context. *In situ* techniques such as in-cell NMR [10–12], fluorescence microscopy [13,14] or cryo-electron tomography [15,16] are developing quickly but still only target individual proteins or protein complexes of interest.

Crosslinking mass spectrometry (CLMS) is a technique that can provide *in situ* middle-resolution structural information for individual multi-protein complexes and can be scaled up to more complex samples such as entire organelles [17] or bacterial cells [18–20]. Distance restraints are generated by identifying which residues were crosslinked in a protein or between two interacting proteins and considering the length of the most extended conformation of the crosslinking reagent. Until recently, complex biological samples could only be tackled by the use of crosslinkers that cleave in the mass spectrometer [19–23] or by the use of isotope-labelled crosslinker which create a special isotope pattern to aide identifying crosslinked peptides [24,25]. Two recent studies investigated murine mitochondria using MS-cleavable crosslinkers and reported 1,876 [22] and 2,779 [23] crosslinked residue pairs (excluding ambiguous crosslinks, where one of the crosslinked peptides could have come from a number of proteins), respectively. These studies focused on the discovery of protein-protein interactions and partially on *in situ* protein structure analysis while possible gains of systematic analysis of protein flexibility have been less explored.

Here, we use *in situ* CLMS to analyse protein structures in human mitochondria. Our crosslink-derived distance restraints combined with high-throughput comparative protein modelling reveal protein interactions and protein flexibility in their native environment. Due to the experimental error associated with single crosslinks we focus our analysis on systematic conflicts between structural models and our *in situ* distance restraints. This critically depends on data density, for which we designed a workflow around a standard crosslinker that combines sequential protein digestion [26], orthogonal peptide fractionation methods, and a decision-tree-based MS acquisition strategy [27]. Importantly, our workflow demonstrates how analysis of complex systems with non-MS-cleavable crosslinkers, including oxidative crosslinkers [28], photoactivatable amino acids [29,30] and photoactivatable crosslinkers [31,32], is now possible and which types of insights that such data add to our understanding of protein structures *in situ*.

Experimental Section

Reagents. Disuccinimidyl suberate (DSS) and Trypsin were purchased from Thermo Scientific Pierce (Rockford, IL). The proteases GluC, Chymotrypsin and AspN were purchased from Promega (Madison, WI).

Cell culture and preparation of human mitochondria. K-562 cells (DSMZ, Cat# ACC-10) were grown at 37 °C under a humidified atmosphere containing 5% CO_2 in RPMI 1640 containing 10% fetal bovine serum. 400 million K-562 cells were collected by centrifugation and washed twice with phosphate buffered saline (PBS). Cell lysis and mitochondria preparation was performed using a protocol adapted from Clayton and Shadel [33]. Briefly, cell lysis was carried out in 5.5 mL ice-cold RSB hypotonic buffer (10 mM HEPES pH 7.5, 10 mM NaCl, 1.5 mM $MgCl_2$) using dounce homogenization. 4 mL ice-cold 2.5 x MS homogenization buffer (12.5 mM HEPES pH 7.5, 525 mM mannitol, 175 mM sucrose, 2.5 mM EDTA) was added to obtain an isotonic solution. To clarify the cell lysate, it was centrifuged three times at

1,300 x g (5 min, 4 °C). The mitochondria were pelleted by centrifugation at 12,360 x g (15 min, 4 °C) and washed once with 5 mL ice-cold 1x MS homogenization buffer. The isolated mitochondria were resuspended in 20 mM sodium phosphate pH 8.0, 150 mM NaCl. The protein concentration was estimated via Bradford Assay (BioRad). Aliquots of isolated mitochondria were frozen in liquid nitrogen and stored at -80 °C.

Crosslinking reaction, tryptic in-solution-digestion and peptide purification. Isolated mitochondria were washed twice in ice-cold PBS and pelleted at 16,000 x g (5 min at 4 °C). 2 mg proteins were chemically crosslinked using 0.225 mM DSS in DMSO which equals a protein-to-crosslinker ratio of 12:1 at 1 mg/mL protein concentration in 2x1 mL in 4x 500 µL crosslinking reaction. Note that a DSS concentration optimization experiment was performed beforehand to find the proximate saturation point of DSS-to-mitochondria ratio by SDS-PAGE and CLMS (further details see supporting material, Fig. S1). After 40 min incubation at 25 °C and gentle agitation, crosslinking reaction was quenched by adding ammonium bicarbonate (ABC) to a final concentration of 50 mM (15 min at 25 °C). Samples were evaporated completely to minimize the volume for tryptic digestion. 2 mg of crosslinked mitochondria were denatured using 6 M urea, 2 M thiourea in 50 mM ABC and reduced with 5 mM dithiothreitol (DTT) 20 min at 50 °C. To alkylate reduced disulfide bonds, 15 mM iodoacetamide (IAA) was added and incubated 30 min at 25 °C in the dark. After diluting with 50 mM ABC to a final concentration of 2 M urea/thiourea, trypsin was added at an enzyme-to-substrate ratio of 1:50 and incubated overnight at 37 °C, with gentle agitation. The in-solution-digestion was stopped by adding 10% (v/v) trifluoroacetic acid (TFA) until pH ≤ 2. Peptide desalting and purification was performed using Empore™ Solid Phase Extraction Cartridges C18-SD according to manufacturer's protocol. Afterwards, the sample was divided in four portions of 500 µg tryptic peptides each.

Fractionation of Peptides by Strong Cation Exchange Chromatography. The tryptic peptides were fractionated using strong cation exchange chromatography (SCX, Fig. 1A) as previously described [34]. In our workflow, four aliquots of each 500 µg peptides were dried in a

vacuum concentrator and resuspended in 105 µL SCX buffer A (20 mM monopotassium phosphate pH 2.7, 30% (v/v) acetonitrile (ACN)). 100 µL peptide samples were loaded onto a PolyLC Polysulfoethyl A™ 100 x 2.1 mm, 3 µm, 300 Å column operated by Shimadzu HPLC system. Peptides were fractionated by increasing the salt concentration (SCX buffer B: 20 mM monopotassium phosphate pH 2.7, 30% (v/v) ACN, 500 mM KCl) with following settings: constant flow rate of 0.2 mL/min, 0% B (0-5 min), 0-5% B (5-10 min), 5-20% B (10-14 min), 20-60% B (14-18 min), 60-70% B (18-21 min), 70-100% B (21-25 min). Two minute fractions were collected and seven selected fractions were partially pooled and evaporated completely, resulting in a total of 5 SCX fractions (14+15, 16, 17, 18 and 19+20 like shown in Fig. S2).

Sequential Digestion (SD) and Size Exclusion Chromatography (SEC). Three of the four 500 µg peptide samples were sequentially digested using a second protease after SCX fractionation (Fig. 1A) as previously described in Mendes et al. [26]. The protease amounts added were adjusted to the peptide content of each SCX fraction. The selected SCX fractions were resuspended in 50 µL 50 mM ABC and digested using either GluC (1:50 protease-to-substrate ratio), Chymotrypsin (1:50) or AspN (1:100). After overnight incubation (Chymotrypsin at 25 °C, GluC and AspN at 37 °C), the protease digestion was stopped using 10% (v/v) TFA. Evaporated sequential digested or tryptic digested samples were resuspended in 40 µL SEC buffer (30% (v/v) ACN, 0.1% (v/v) TFA) and fractionated using SEC as previously described [35]. In our workflow, peptides were fractionated using a Superdex Peptide 3.2/300 column (GE Healthcare) operated by Shimadzu HPLC system at a flow rate of 0.05 mL/min in a 60 min isocratic gradient with SEC buffer. Two minute fractions were collected and, depending on the sample amount, two up to six early eluting SEC fractions were selected (Fig. S2). Due to the expectation that crosslinked peptides are overall larger than linear peptides, we selected only early SEC fractions for MS acquisitions. This entire workflow resulted in 88 different SCX-SD-SEC fractions which were evaporated completely and resuspended in 4 µL 0.1% (v/v) FA.

LC-MS/MS acquisition. A total of 110 MS runs were analysed as previously described [27] using an UltiMate 3000 Nano LC system coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, USA). SCX-SD-SEC fractions with large sample amounts were injected as technical duplicates. Briefly, mobile phase A contained 0.1% (v/v) FA in water and mobile phase B 80% (v/v) ACN and 0.1% (v/v) FA in water. Fractionated peptides were injected onto a 500 mm C-18 EasySpray column (75 µm ID, 2 µm particles, 100 Å pore size) and separated using a constant flow rate of 250 nL/min. Depending on the sample amount per fraction, a linear gradient from 4-40% mobile phase B was employed for either 60 min or 139 min for peptide elution. MS1 spectra were acquired at 120,000 resolution in the orbitrap with AGC target of 2×10^5 ions and a maximum injection time of 50 ms. For fragmentation precursor ions with charge states 3-8 and an intensity higher than 5×10^4 were isolated using an isolation window of 1.4 m/z (AGC target $1\text{-}5 \times 10^4$, 60 ms max. injection time). Depending on the charge state and the m/z ratio, precursor ions were fragmented with energies based on the optimized data-dependent decision tree using HCD/EThcD fragmentation [27]. MS2 spectra were recorded at 30,000 resolution in the orbitrap.

Identification and validation of crosslinked peptides. MS raw data were converted to mgf format using msconvert, including a peak filter for the 20 most abundant peaks per 100 m/z window for further data analysis (Fig. 1A). Resulting peak files were analysed by Xi (version 1.6.731 [26], using the following settings: MS tolerance 6 ppm, MS2 tolerance 20 ppm, potential missing monoisotopic peaks 3 [36], enzyme dependent on respective single or sequential digestion (trypsin or trypsin + AspN, trypsin + GluC, trypsin + chymotrypsin), fixed modification carbamidomethylation of cysteine, variable modification oxidation on methionine, losses $-CH_3SOH$, $-H_2O$ and $-NH_3$, crosslinker BS3 (mass equivalent in crosslinked state, mass modification 138.06807 Da) with variable crosslinker modifications (“BS3-NH2” 155.0946 Da, “BS3-OH” 156.0786 Da). The DSS crosslinker was assumed to react primarily with lysine residues, but also with serine, threonine, tyrosine or the protein N-terminus. Besides precursor-ions were the corresponding b- and y-fragment ions searched for HCD

fragmentation; for EThcD b-, c-, y- and z-fragment ions were considered. Obtained peptide spectra were matched to a database constructed either from the MitoCarta 2.0 database of annotated human mitochondrial proteins (1,157 protein IDs [6], Tab. S1B) or including the most abundant proteins in all crosslinked and SCX-SD-SEC fractionated samples (1,118 protein IDs Tab. S1B, see expanded material of linear MaxQuant Search in Tab. S1A). To filter for high confidence data a false discovery rate (FDR) of 5% on link level was applied on the identified crosslinked peptides using xiFDR [37]. Note that unlike other FDR calculations, our xiFDR groups identified peptides into putative self- or PPI-links to avoid an accumulation of false positives for self-links. Crosslinks within one protein were calculated using following settings: pre-filter crosslinks only, 5 amino acids as minimum peptide length. Crosslinks between two different proteins were analysed with the following parameters: pre-filter crosslinks only, delta score 0.5, minimum number of fragments per peptide 5, with 8 amino acids as minimum peptide length.

Mitochondrial protein localisation. The known localizations of 915 mitochondrial proteins were assigned according to MitoCarta and UniProt subcellular location information (see Tab. S1B). Remaining mitochondrial proteins from MitoCarta were annotated as “other mitochondrial localisation”.

Crosslink assessment using models from PDB. We investigated crosslinks by mapping the residue pairs to all available PDB structures. For crosslinks within the same protein, we mapped the crosslinks on available monomeric structures and, where applicable, also on homomultimeric structures. For some proteins there are several PDB structures or comparative models available, in which we mapped our self-links to the shortest distance in any given structure. Then, for each crosslinked residue pair we calculated the Euclidean distance between the C_α atoms within the PDB structures. We consider a crosslink in agreement with the PDB model if the C_α-C_α distance is smaller or equal to 30 Å [38]. If not, we consider a crosslink to be a long-distance link. We use the SIFTS database [39] to map the canonical UniProt sequences in our search database to available PDB structures. Note that

for a unique, canonical UniProt [40] sequence, there might be multiple PDB structures available. In this case, we calculate the distances for all PDB structures and take the shortest, including homooligomeric interfaces.

Protein structure modelling. We performed comparative modelling on 363 proteins with unknown structure. The modelling procedure consists of four steps: First, we generate a sequence profile of the target sequence by searching for homologous sequences with HHblits 3.0.0 [41]. Second, used the profiles to search the PDB70 database from February 2017 using HHSearch [42]. We accept a template for a given protein sequence, if the negative logarithm of the HHSearch had a p-value ≥ 6.5 [43], which corresponds to the threshold for a remote structure. If all templates for a protein do not satisfy this criterion, we do not model the structure because no reliable templates can be identified by HHsearch. Third, we used MODELER 9.12 [44] to generate fifty comparative models for each protein. Fourth, we used PROSA [45] to select the top scoring comparative model for each protein. We used normal mode analysis (NMA) to model dynamics by using the web-based elNémo software [46,47] and ProDy [48]. During modelling, we use the template quality as a proxy to measure the quality of the resulting comparative models. HHsearch found template hits for 654 proteins for which we also identified crosslinks, 363 of which had no experimental structure in humans. Swiss-model [49] was used for the modelling of the ATP synthase as well as for the complex III of the OXPHOS complex.

Structure visualisation and protein docking. Structures were visualized with UCSF Chimera [50] and PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC. During highlighting crosslinked amino acids in the Hsp60 protein complex (Fig. 3B/C), the crosslinked K551 was not present in the PDB structure, therefore the neighbouring amino acids were modelled using MODELLER [44]. For protein docking, we used the HADDOCK web service with default parameters [51,52]. Crosslinks between domains were set as unambiguous distance restraints with an upper limit of 30 Å in docking calculations. Centre of mass restraints was

enabled. To account for the peptide sequence between the domains, we imposed an upper limit of 35 Å.

Results and Discussion

Crosslinking of human mitochondria using a non-cleavable crosslinker

Human mitochondria (K-562 cells) were crosslinked using the homobifunctional and membrane permeable reagent disuccinimidyl suberate (DSS) (Fig. 1A). Following tryptic digestion, peptides were fractionated using strong cation-exchange chromatography (SCX) into five fractions with an enrichment of crosslinked peptides in higher salt fractions [24,53] (Fig. S2). The individual fractions were then subjected to our novel sequential digestion protocol (see Experimental Section for details, Fig. 1A) [26]. The second digestion step preferentially shortens large and thus difficult-to-observe peptides, due to the proteases having reduced cleavage efficiency for shorter peptides [26,54]. All fractions were then subjected to size-exclusion chromatography (SEC) to further enrich for crosslinked peptides (Fig. S2B) [26,35,54]. Only early SEC-fractions, those enriched for crosslinked peptides, were analysed by LC-MS/MS using a data-dependent decision tree of optimised fragmentation energies for crosslinked peptides [27]. The database of protein sequences for crosslink search was generated by combining the most abundant 1,000 proteins (Tab. S1A) in our mitochondrial preparations with all proteins listed in MitoCarta (total proteins 1,660, Tab. S1B).

In total, we identified 12,664 unique crosslinked peptide pairs (excluding ambiguous crosslinks, Tab. S2A) which correspond to 5,518 unique residue pairs in 792 proteins (5% link-level FDR [37]). The majority of the proteins and protein-protein contacts were identified by multiple residue pairs (Fig. 1B/C). Of these, 5,366 are putative self-links (Molecular Interactions Controlled Vocabulary ID: 0898, Tab. S2B, from here referred to as “self-links” for simplicity). Self-links fall either within one or between two copies of the same protein. Some proteins such as malate dehydrogenase (MDHM), mitochondrial stress 70 protein (GRP75)

and 60 kDa heat shock protein (Hsp60) were covered by more than 100 residue pairs (Fig. 1B), which suggests the presence of abundant structural information in our data at least for some proteins. Overall 4,034 self-links mapped to 513 (44% of 1,157) proteins of MitoCarta (Fig. 1D). The majority of identified links were identified on mitochondrial matrix proteins consistent with DSS crosslinker passing through both mitochondrial membranes (54%, Fig. 1D), although we cannot exclude the presence of fractured or lysed mitochondria. Furthermore, we identified 1,335 crosslinks in 255 proteins that are not included in MitoCarta. According to UniProt, some of these proteins are localised in mitochondria. However, the majority of these proteins are assigned to cytosolic cellular functions or belong to the endoplasmic reticulum which is connected to the mitochondria and likely constitute background of our purification. Nevertheless, our human mitochondria preparation was highly enriched. As one would expect, we identified crosslinks exclusively in the more abundant proteins (Fig. S3) while crosslinks between proteins displayed an even higher bias towards highly abundant proteins. This underpins the general challenge of detecting crosslinked peptides. This also supports our decoy-based FDR-approach for error assessment as random false identifications should not show an abundance bias.

Despite our departure from non-MS-cleavable crosslinkers, we identified more crosslinks than previous studies using MS-cleavable crosslinkers (Fig. S5A). There is a set of possible contributing factors: We countered some of the disadvantages of standard crosslinkers by optimised data acquisition [27] and breakdown of the combinatorial search space [55]. In contrast to others, our study employed sequential digestions [26] which boost our number of identifications up to 65% by shortening the average peptide length from 33 amino acids for crosslinked tryptic peptides to 22-24 amino acids for crosslinked sequentially digested peptides (Fig. S4A/B). However, tryptic data contributed 4,481 peptide pairs (Fig. S4C) which still compares favourably to previous analyses of mitochondria which yielded in total 2,427 [22] and 2,779 peptide pairs [23]. Note that although NHS-esters preferably crosslink primary amines such as found in lysine side chains, there is a known side reaction with the

hydroxyl groups of S/T/Y residues [56,57]. These crosslinks were not considered in the previously published studies, but they contributed to 3,066 crosslinked peptide pairs (33%) in our full dataset (Fig. S4D/E). Moreover, monoisotopic peak correction during database search [36] make up to 40% peptide spectrum matches in our full dataset (Fig. S4F). However, a direct comparison of all three studies is hampered by many parameters that differ between them, including sample origin, digestion method, fractionation methods, fractionation depths, acquisition method and time, data analysis software and finally FDR estimation with various filter settings and grouping of PPI- and self-links [37].

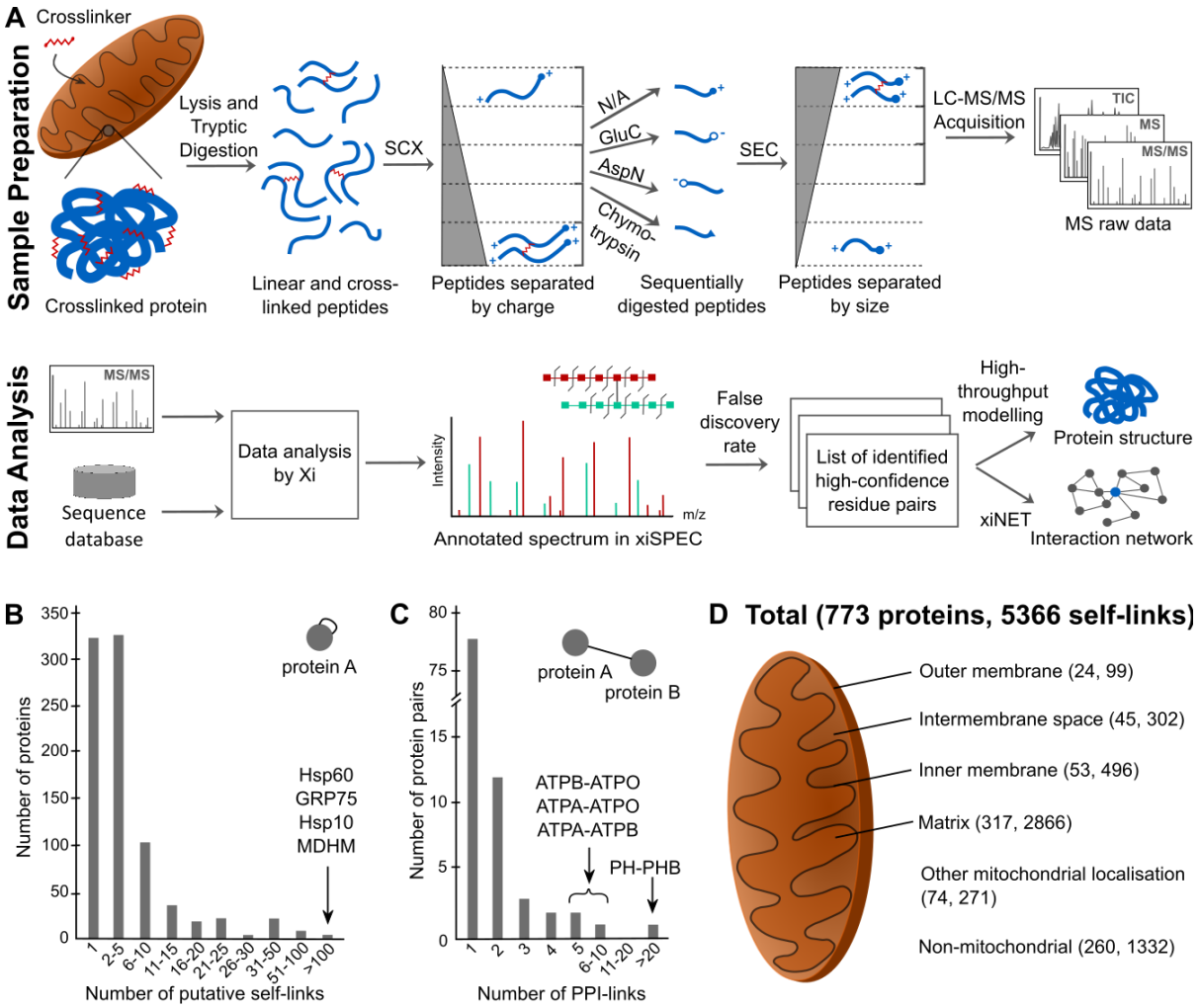

Fig. 1: Workflow, data density and quality of crosslinking mass spectrometry analysis in human mitochondria. (A) Overview of crosslinking pipeline in human mitochondria. Sample preparation (upper panel): Isolated mitochondria were crosslinked using the membrane-permeable crosslinker disuccinimidyl suberate (DSS). Proteins were digested with trypsin and the resulting peptides were fractionated by strong cationic exchange

chromatography (SCX). Each fraction was then subjected to size exclusion chromatography (SEC) which enriches for crosslinked peptides in early fractions. SEC was conducted either directly or following an additional digestion step by either GluC, AspN or chymotrypsin, which preferentially cleaves large peptides to enhance their detection during the subsequent mass spectrometric analysis [26,54]. Data analysis (lower panel): The acquired MS/MS spectra were searched against a sequence database using Xi [26]. Crosslinks were filtered to 5% false discovery rate using xiFDR [37] and used to analyse protein-protein interactions in xiNET [58] and for protein structure modelling. (B) The majority of proteins detected with putative self-links are seen with multiple crosslinks. Stress 70 protein (GRP75), malate dehydrogenase (MDHM) and 60 kDa heat shock protein (Hsp60) have more than 100 self-links. (C) The majority of protein pairs identified with crosslinks are based on a single PPI-link. Protein-protein interactions between ATP synthase subunits (ATPA, ATPB, ATPO) and prohibitin – prohibitin 2 (PHB-PHB2) are characterised by up to 20 unique PPI-links. (D) Localisation of identified residue pairs of self-links within the human mitochondrion.

CLMS data reveals conformations adopted by proteins in situ

We compared our self-links against experimental structures deposited in the PDB or, where none were available, to comparative models based on structures from other species (see Experimental Section). 2,215 (41.3% of 5,366) crosslinks mapped on 343 proteins with human PDB structures (green subset in Fig. S6A/C, Tab. S4A) and furthermore 1,290 crosslinks on 256 proteins with comparative models (blue subset in Fig. S6A/D, Tab. S5A). Focussing on monomeric PDB structures, 219 crosslinks (9.9%, Fig. S6B) surpassed 30 Å Cα-Cα distance, an empiric upper boundary for DSS crosslinking that is also supported by molecular dynamics simulations [38]. By considering known homomultimerisation, this reduced to 129 long-distance self-links in PDB entries (5.8%, Fig. S6B/C). Thus, considering homomultimeric states resolved conflicts for 90 links. At least 66 of the remaining 129 long-distance self-links in 47 PDB entries will be rationalised below in the context of conformation changes, further reducing the apparent conflict between our self-links and PDB data to below 3% (Fig. S6B, Tab. S4B). Additionally, we identified 68 crosslinks with zero sequence separation (Tab. S2B) which cannot stem from the same protein molecule. These self-links may indicate homomultimerisation, but may also be artefacts due to non-covalent association of the

peptides during mass spectrometric measurement [59] and thus were excluded from our structural analyses.

We investigated clusters of long-distance links to see whether they may reveal novel structural states *in situ*. As mentioned, homomultimerisation resolved long-distance links. Some of these were clustered, for example in the case of the beta subunit of the methylcrotonoyl-CoA carboxylase as part of the MCCC complex. 15 self-links match the monomeric structure while three were in conflict with it (Fig. S7A left panel). Using the oligomeric orthologous from *Pseudomonas aeruginosa* (PDB 3U9S) as template for modelling harmonised these conflicts (Fig. S7A right structure), consistent with a homooligomeric structure also of the human MCCC2 complex.

Furthermore, we found several cases of clustered conflicts which indicate protein flexibility *in situ*. The mitochondrial elongation factor Tu had six long-distance links (out of 57 self-links) which connect from different parts of the protein to the beta sheet domain (shown at the bottom in Fig. S7B, left panel). A normal mode analysis of our comparative model suggests a domain movement towards the core structure (indicated with an arrow in Fig. S7B, right panel) that reduces all long-distance links. We find an analogous protein flexibility in the mitochondrial OXPHOS supercomplex (complex $I_1III_2IV_1$) which is critical for ATP production in mitochondria. Overall, we see an excellent agreement with previous structural data (PDB 5XTH, Fig. S7C), but 20 out of 196 distance restraints (10.2%) exceed 30 Å C_α-C_α distance. Seven long-distance links clustered in complex I, involving the proteins NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7 (NDUA7), NADH dehydrogenase [ubiquinone] iron-sulfur protein 2 and 3 (NDUS2, NDUS3). Especially NDUA7 consists of extensive unstructured protein segments (Fig. S7C) which contributes to protein flexibility in this region of complex I. Extending *in situ* structural analysis to another OXPHOS complex, we visualised distance restraints also in ATP synthase (complex V). In the absence of a human structure, we mapped human protein sequences into the bovine structure, which is available in different states of the ATP production cycle [60]. Resulting models and crosslinks were in excellent

agreement (187 out of 199 fall below 30 Å, Fig. S7D). Crosslinks fell within each of the major extra membrane domains of the ATP synthase (rotor, peripheral stator and $\alpha_3\beta_3$ core subunit). The 12 distance restraints that exceed our 30 Å cut-off did not cluster and were thus not used to propose conformational changes.

The two structurally solved domains of Hsp70 (PDB 4KBO, 3N8E) covered 60 of the identified 134 self-links and could be arranged using the full-length model of the *E. coli* orthologue (PDB 2KHO, grey structure in Fig. 2A upper panel). However, crosslink data disagreed with the resulting interface of both domains (highlighted in orange, Fig. 2A) for which also in *E. coli* some flexibility has been reported [61]. Docking the domains using crosslink restraints in HADDOCK [51,52] resolved many conflicting crosslinks (Fig. 2A lower panel). This previously undescribed arrangement proposed by CLMS might occur during protein regulation. In fact, crosslinking also captured the regulatory mitochondrial GrpE protein homolog 1 [62] in the substrate binding domain of Hsp70 (Tab. S6). Our CLMS data therefore suggest that negative regulation by GrpE may require a dramatic dynamic process of both Hsp70 domains (as indicated by an arrow in Fig. 2A lower panel).

The most striking conflict with established structures was observed for the 60 kDa heat shock protein (Hsp60). We identified 291 self-links, including 61 long-distance crosslinks, almost all of which indicate a compression of the heptameric Hsp60 ring (PDB 4PJ1, Fig. 2B). The orthologous chaperonin GroEL/GroES system in *E. coli* has flexible apical domains that can be attributed to the ability of chaperonin to bind different substrates or to the involvement of the apical domains in substrate unfolding processes [63–65]. In comparison, the human chaperonin structure shows more intense asymmetric movements within the Hsp60 ring subunits which were previously suggested not to be concerted [66]. In contrast, our *in situ* CLMS data show clustered long-distance restraints with anchor points on the opposite side of the heptameric Hsp60 ring structure (Fig. 2B). This indicates a compression-like movement of Hsp60 *in situ* (indicated with an arrow in Fig. 2B). Therefore we mapped also our distance restraints to the GroEL structure (PDB 4AAQ) which has a narrow structure (Fig. 2C). This

solved nearly half of the conflicting distance restraints. The remaining conflicts suggest an even higher degree of protein flexibility. Consequently, substrate binding and/or unfolding might require directed movements within the human chaperonin ring similar to but possibly exceeding those described for the ATP-dependent *E.coli* GroEL/GroES system [64,65].

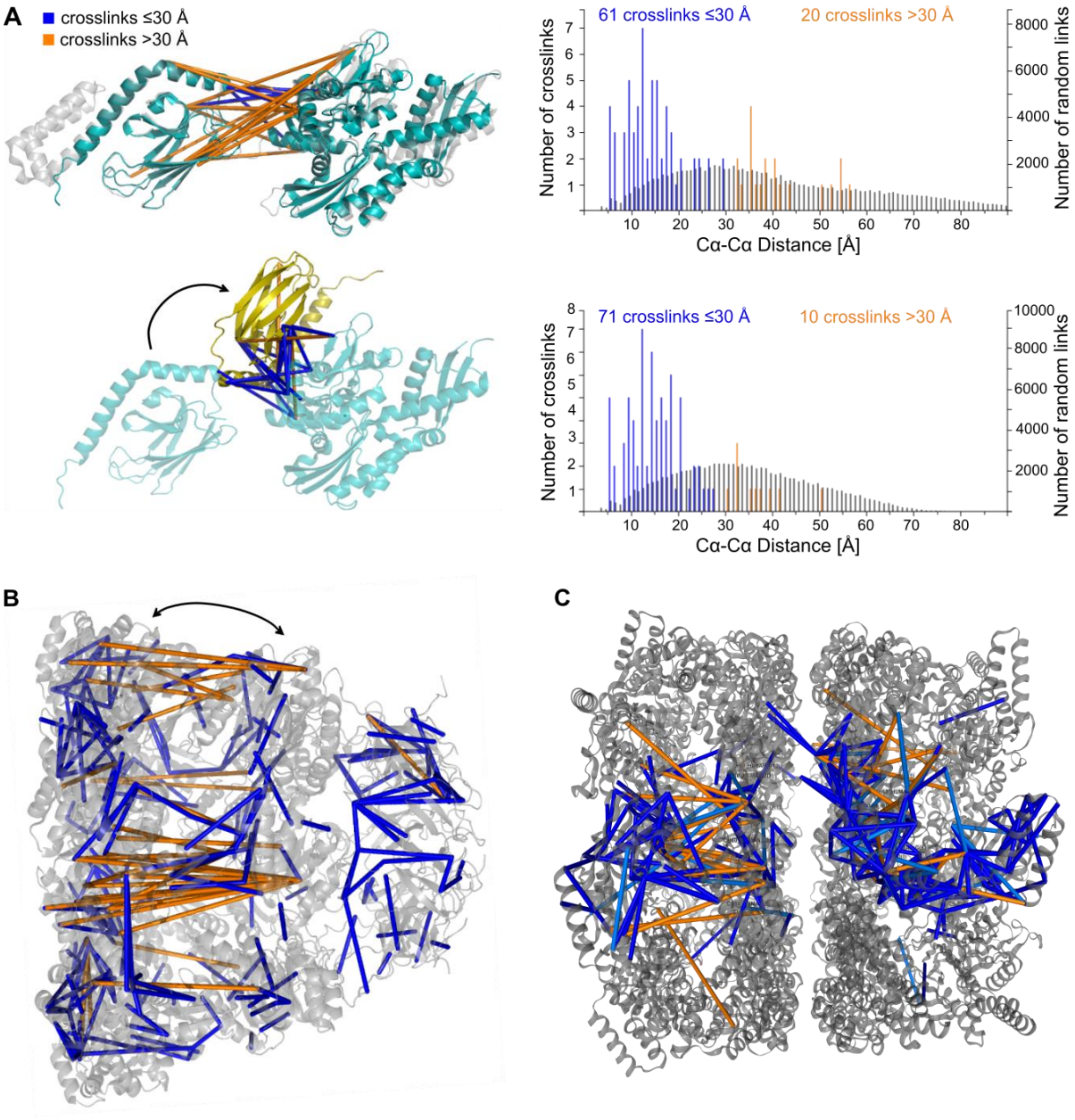

Fig. 2: *In situ* determined self-links contain structural information. (A) Full-length modelling of stress 70 protein. Positioning two human Hsp70 domains structures (PDBs 4KBO/4N8E shown in cyan in upper structure) using Hsp70 structure in *E. coli* (PDB 2KHO shown in grey) as a template. The majority of conflict restraints are at the domain interface, which are indicated in the structure. Docking with the CLMS restraints resolved most of the

long-distance links at the interface and suggests alternative domain arrangement of Hsp70 (lower structure, shown in yellow). Histograms show the length distribution of all distance restraints on these protein structures before and after docking. (B) *In situ* flexibility within the mitochondrial chaperonin complex. The PDB structure 4PJ1 (shown in grey) portrays the heptameric 60 and 10 kDa heat shock protein complex. Identified crosslinks were matched to a single ring structure at the shortest distance between the crosslinked residues. (C) Assessment of human crosslinks in the context of an *E. coli* homologue. When mapping the human *in situ* CLMS distance restraints on GroEL (PDB 4AAQ) half of the conflicting restraints were resolved.

In situ protein-protein interactions revealed by CLMS

In addition to self-links, we identified 152 PPI-links (protein-protein interaction links between two proteins, Tab. S2C). Our interaction network comprised 134 crosslinks between mitochondrial proteins (Fig. 3A). This included known interactors such as the respiratory chain complexes, ATP synthase, mitochondrial heat shock proteins and prohibitin – prohibitin 2 interaction. We also identified 26 protein-protein interactions which are not yet annotated for human mitochondria in STRING or BioGrid databases (highlighted red in Fig. 3A). Some could be explained as possible substrates for mitochondrial heat shock proteins (see Hsp60 results described below). Others describe interactions between subunits of the respiratory chain complexes I, III, IV and V (ATP synthase) like the interaction between NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5 and 7 (NDUA5 and NDUA7) which lacked an experimental evidence in humans so far but is known in putative homologs. Notably, multiple crosslinks (Tab. S2C) support the interaction between ATPase family AAA domain-containing protein 3A and 3B (ATD3A and ATD3B). Using loss- and gain-of-function approaches, Merle *et al.* [67] showed that the association of ATD3B with the ubiquitous ATD3A protein negatively regulates the interaction of ATD3A with matrix nucleoid complexes and contributes to mitochondrial homeostasis and metabolism specific in embryonic stem cells. We here found evidence for these heterodimers also in mature K-562 cells.

We identified a much lower fraction of PPI-links (2.8% of all crosslinks, Fig. S5A), than the previous studies Schweppe *et al.* (29%) and Liu *et al.* (64%). As a plausible contributing

factor, we investigated different ways of FDR calculation employed by these studies. In contrast to Schweppe *et al.*[22] and Liu *et al.*[23], we separate PPI- and self-links for FDR analysis, due to a large prior probability that self-links are correct [37]. If we do not separate these for FDR estimation, we see a substantial increase in PPI-links (on our tryptic subset, 16% up from 3.3%, Fig. S5B). Unfortunately, many of these additional PPI-links are likely false, we gain 53 PPI-links but also 22 PPI-link decoys, i.e. FDR 42%. Also, self-links decrease 3.4-fold, from 2620 to 767. The need for separating PPI- and self-links for FDR analysis is further supported by all three studies seeing the majority of self-links supported by multiple peptide pairs (Fig. S5C). However, only after separate FDR estimation is this also the case for PPI-links (Fig. S5D). Taken together, this reveals a large dependency of PPI-links on the FDR method and suggests that the field needs to find a standardised and carefully tested agreement here.

Nevertheless, different mitochondrial study approaches can corroborate each other by independently identifying novel interactions. We compared protein-protein interactions identified in our study to orthologous interactions found in murine mitochondria by Schweppe *et al.*[22] and Liu *et al.*[23]. Here, 8 out of 14 protein-protein interactions are revealed by a single PPI-link (our least statistically confident protein-pairs) in our data were also found in mouse (Tab. S3). For example, ATP synthase subunits **α** (ATPA) and d (ATP5H) were observed with multiple links in both mouse studies. Additionally, this approach also supported 6 (blue lines in Fig. 3A, Tab. S3) out of 26 protein-protein interactions that were not reported for human mitochondria in STRING or BioGrid (red lines in Fig. 3A).

The presence of an Hsp60 structure allowed a closer look at the 17 PPI-links involving Hsp60 and other proteins. Importantly, 5 out of the 11 Hsp60 residues involved in these links located in the substrate channel at the inside of the Hsp60 ring structure, indicating that the proteins linked to those are likely substrates (Fig. 3B, Tab. S6). This includes a kinase, the mitochondrial glycerol kinase (GLPK) and a phosphatase, mitochondrial phosphatidylglycerophosphatase and protein-tyrosine phosphatase 1 (PTPM1). Previous studies have shown that the *E. coli* GroEL/GroES system folds a wide spectrum of proteins

including certain kinases and phosphatases [68,69] but chaperonin substrates in the human system remain mostly unknown.

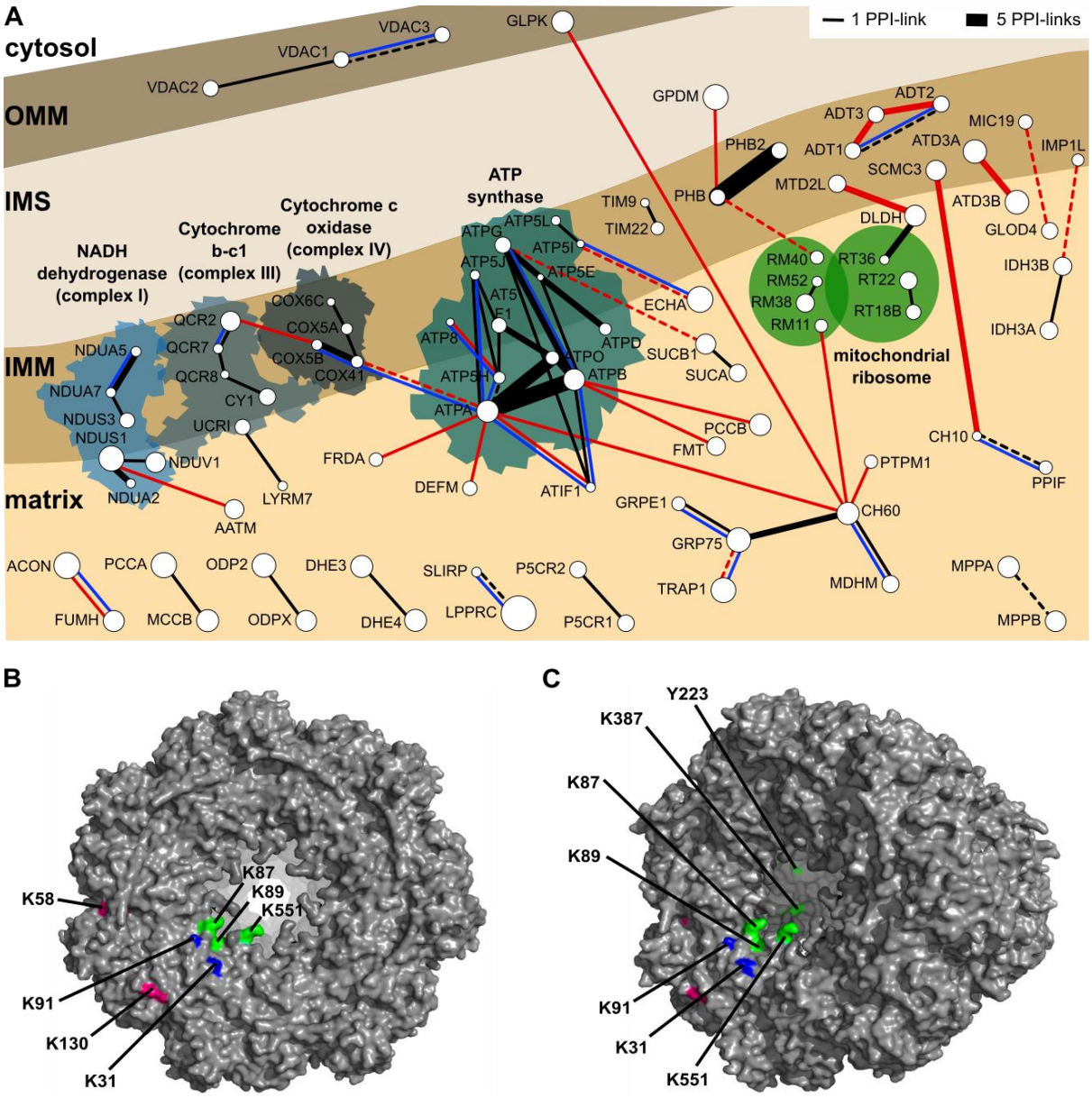

Fig. 3: Protein-protein interaction analysis by CLMS. (A) Interaction network in human mitochondria. White circles represent proteins for which PPI-links were identified and lines illustrate these interactions. Thickness of lines scales with the number of PPI-links for each interaction. Lines are dashed when only one crosslink was detected. Lines are coloured according to interactions found in STRING or BioGrid database (black) or not (red). Additional blue lines indicate that this particular protein-protein interaction was also identified by Schweppe *et al.* and/or Liu *et al.* (Tab. S3). The most dense interaction network in human mitochondria was observed in the complexes of the oxidative phosphorylation, mitochondrial heat shock proteins and prohibitin. OMM: outer

mitochondrial membrane, IMS: intermembrane space, IMM: inner mitochondrial membrane. (B/C) Crosslinked amino acids in the 60 kDa heat shock protein complex, chain A. Residues being highlighted in green are located in the substrate channel, at the inside of the Hsp60 barrel. For example, Y223 crosslinks to GLPK and K387 connects to PTPM1. Furthermore, crosslink sites localise to the interface of the two heptameric Hsp60 ring structures, including K87/89/551 (highlighted in green) and the K31/91 (highlighted in blue). These amino acids crosslink to MDHM or Hsp70 (see Tab. S6 for further information). Crosslinked residues at the outside of the barrel are coloured in pink.

Conclusion

In summary, these results support the use of CLMS as an *in situ* structural analysis method, to gain new insights into multimerisation and protein flexibility occurring *in situ*. Most of the proteome-wide CLMS studies to date focus on protein-protein interactions. We add a systematic view on clustered conflicts of long-distance crosslinks within proteins. We show that large-scale CLMS, even by using a standard non-cleavable crosslinker, generates sufficient data to start informing protein structure analysis across an entire cellular organelle. This workflow releases constraints for novel crosslinker designs and opens complex mixture CLMS to other crosslinker chemistries such as oxidative crosslinking or use of photoactivatable crosslinkers and amino acids. By further maturation of proteome-wide CLMS analysis, there will be more over-length conflicts which can be explained in a biological context. This will extend our structural knowledge in a unique way, being complementary to traditional structure elucidation methods.

Associated Content

Supporting Information. The following information is available free of charge at ACS website: http://pubs.acs.org. Figure S1: DSS concentration optimization experiment. Figure S2: Coupled SCX and SEC fractionation of tryptic peptides. Figure S3: Protein abundance of linear identifications in comparison to proteins with identified self- and PPI-links. Figure S4: Detailed peptide pair analysis of our identified crosslinks. Figure S5: Comparison of three different CLMS mitochondria studies. Figure S6: Assessment of PDB structures and comparative models using *in situ* CLMS data. Figure S7: *In situ* determined self-links contain structural information.

Table S1A: MaxQuant Search output table of crosslinked and fractionated mitochondria samples originated from all sequential digestion datasets. Table S1B: Protein IDs of the most abundant proteins in crosslinked and fractionated sample and protein IDs of MitoCarta. Table S2A: List of all identified unique peptide pairs after 5% FDR calculation on link-level. Table S2B: List of all identified putative self-crosslinks after 5% FDR calculation on link-level. Table S2C: List of all identified PPI-links after 5% FDR calculation on link-level. Table S3: Auxiliary crosslink identifications in human and murine mitochondrial proteins. Table S4A: Self-links mapped on available human monomeric PDB structures. Table S4B: Candidates with systematic mismatches within monomeric PDB models. Table S5A: Self-links mapped on comparative models. Table S5B: Candidates for systematic mismatches in comparative models. Table S6: List of identified crosslinks that indicate protein-protein interaction to human heat shock proteins like Hsp60.

Data availability. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [70] partner repository with the dataset identifier PXD014675. Other data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Author information

Corresponding Author. *E-Mail: juri.rappsilber@tu-berlin.de

ORCID. Petra S. J. Ryl: 0000-0003-3016-3874, Swantje Lenz: 0000-0002-8839-5371, Lutz Fischer: 0000-0003-4978-0864, Lisa Budzinski: 0000-0003-4120-6032, Marchel Stuiver: 0000-0003-3437-4468, Marta M. L. Mendes: 0000-0002-3528-7123, Ludwig Sinn: 0000-0003-4692-0681, Francis J. O'Reilly: 0000-0001-9258-0150 and Juri Rappsilber: 0000-0001-5999-1310.

Authors Contribution. P.R. and J.R. developed the study; M.S. performed cell culture and mitochondria isolation; P.R. and L.B. performed sample preparation and chromatographic fractionation in cooperation with L.S.; P.R. and M.M. performed LC-MS analysis; P.R., M.B.S., L.F., S.L., F.O'R. and J.R. performed data analysis; M.B.S., S.L. and P.R. performed structural modelling; P.R., M.B.S., F.O'R. and J.R. wrote the article with critical input from all authors.

Conflict of Interest Disclosure. The authors declare no competing financial conflict of interest.

Acknowledgements

We thank Andrea Graziadei for the technical support for protein modelling and helpful discussions. We thank Dr. Philipp Selenko for the opportunity of using the cell culture and lab equipment for mitochondria isolation at the FMP, Berlin. We thank Dr. Janine Kirstein for valuable discussion and comments on our data on heat shock proteins. Furthermore, we thank Dr. James E. Bruce and Dr. Juan Chavez for critically reading the manuscript. This work was supported by the Einstein Foundation, the DFG [RA 2365/4-1, SFB-740 1-5000019-01-TP] and the Wellcome Trust through a Senior Research Fellowship to JR [103139] and a multi-user equipment grant [108504]. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust [203149].

References

1. Duchen, M. R. Mitochondria in health and disease: perspectives on a new mitochondrial biology. *Mol. Aspects Med.* **25**, 365–451 (2004).
2. Cadonic, C., Sabbir, M. G. & Albensi, B. C. Mechanisms of Mitochondrial Dysfunction in Alzheimer's disease. *Mol. Neurobiol.* **53**, 6078–6090 (2016).
3. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat Rev Dis Primers* **2**, 16080 (2016).
4. Vyas, S., Zaganjor, E. & Haigis, M. C. Mitochondria and Cancer. *Cell* **166**, 555–566 (2016).
5. Tahrir, F. G., Langford, D., Amini, S., Mohseni Ahooyi, T. & Khalili, K. Mitochondrial quality control in cardiac cells: Mechanisms and role in cardiac cell injury and disease. *J. Cell. Physiol.* **234**, 8122–8133 (2019).
6. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–7 (2016).
7. Acharya, K. R. & Lloyd, M. D. The advantages and limitations of protein crystal structures. *Trends Pharmacol. Sci.* **26**, 10–14 (2005).
8. Wang, H.-W. & Wang, J.-W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci.* **26**, 32–39 (2017).
9. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85 (2016).
10. Li, H. & Sun, H. In-cell NMR: an emerging approach for monitoring metal-related events in living cells. *Metallomics* **6**, 69–76 (2014).
11. Freedberg, D. I. & Selenko, P. Live cell NMR. *Annu. Rev. Biophys.* **43**, 171–192 (2014).
12. Rhodes, C. J. Magnetic resonance spectroscopy. *Sci. Prog.* **100**, 241–292 (2017).
13. Heintzmann, R. & Huser, T. Super-Resolution Structured Illumination Microscopy. *Chem. Rev.* **117**, 13890–13908 (2017).
14. Li, W., Geng, C. & Liu, B. Visualization of synaptic vesicle dynamics with fluorescence proteins. *Folia Neuropathol.* **56**, 21–29 (2018).
15. Wagner, J., Schaffer, M. & Fernández-Busnadiego, R. Cryo-electron tomography-the cell biology that came in from the cold. *FEBS Lett.* **591**, 2520–2533 (2017).
16. Koning, R. I., Koster, A. J. & Sharp, T. H. Advances in cryo-electron tomography for biology and medicine. *Ann. Anat.* **217**, 82–96 (2018).
17. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nature Structural & Molecular Biology* **25**, 1000–1008 (2018).
18. Zheng, C. *et al.* Cross-linking measurements of in vivo protein complex topologies. *Mol. Cell. Proteomics* **10**, M110.006841 (2011).

19. Weisbrod, C. R. *et al.* In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J. Proteome Res.* **12**, 1569–1579 (2013).
20. Navare, A. T. *et al.* Probing the protein interaction network of Pseudomonas aeruginosa cells by chemical cross-linking mass spectrometry. *Structure* **23**, 762–773 (2015).
21. Chavez, J. D., Weisbrod, C. R., Zheng, C., Eng, J. K. & Bruce, J. E. Protein interactions, post-translational modifications and topologies in human cells. *Mol. Cell. Proteomics* **12**, 1451–1467 (2013).
22. Schweppe, D. K. *et al.* Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1617220114
23. Liu, F., Lössl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol. Cell. Proteomics* **17**, 216–232 (2018).
24. Rinner, O. *et al.* Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318 (2008).
25. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
26. Mendes, M. L. *et al.* An integrated workflow for cross-linking/mass spectrometry. *bioRxiv Archive 355396* doi:10.1101/355396
27. Kolbowski, L., Mendes, M. L. & Rappsilber, J. Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. *Anal. Chem.* **89**, 5311–5318 (2017).
28. Cline, E. N. *et al.* A novel crosslinking protocol stabilizes amyloid β oligomers capable of inducing Alzheimer's-associated pathologies. *J. Neurochem.* **148**, 822–836 (2019).
29. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2**, 261–267 (2005).
30. Lössl, P. *et al.* Analysis of nidogen-1/laminin γ1 interaction by cross-linking, mass spectrometry, and computational modeling reveals multiple binding modes. *PLoS One* **9**, e112886 (2014).
31. Belsom, A., Schneider, M., Fischer, L., Brock, O. & Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Mol. Cell. Proteomics* **15**, 1105–1116 (2016).
32. Tanaka, Y., Bond, M. R. & Kohler, J. J. Photocrosslinkers illuminate interactions in living cells. *Mol. Biosyst.* **4**, 473–480 (2008).
33. Clayton, D. A. & Shadel, G. S. Isolation of mitochondria from cells and tissues. *Cold Spring Harb. Protoc.* **2014**, db.top074542 (2014).
34. Fritzsche, R., Ihling, C. H., Götze, M. & Sinz, A. Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Communications in Mass Spectrometry* **26**, 653–658 (2012).

35. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.* **9**, 120–137 (2014).
36. Lenz, S., Giese, S. H., Fischer, L. & Rappsilber, J. In-Search Assignment of Monoisotopic Peaks Improves the Identification of Cross-Linked Peptides. *Journal of Proteome Research* **17**, 3923–3931 (2018).
37. Fischer, L. & Rappsilber, J. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal. Chem.* **89**, 3829–3833 (2017).
38. Merkley, E. D. *et al.* Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* **23**, 747–759 (2014).
39. Velankar, S. *et al.* SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res.* **41**, D483–9 (2013).
40. Poux, S. *et al.* On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. *Bioinformatics* **33**, 3454–3460 (2017).
41. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).
42. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
43. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences* **110**, 15674–15679 (2013).
44. Webb, B. & Sali, A. Protein Structure Modeling with MODELLER. *Methods Mol. Biol.* **1654**, 39–54 (2017).
45. Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Genetics* **17**, 355–362 (1993).
46. Suhre, K. & Sanejouand, Y.-H. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32**, W610–W614 (2004).
47. Suhre, K. & Sanejouand, Y. H. On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 796–799 (2004).
48. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
49. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
50. Pettersen, E. F., Goddard, T. D. & Huang, C. C. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Comput Chem* **13**, 1605-1612 (2004).
51. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).

52. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
53. Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO Journal* **29**, 717–726 (2010).
54. Dau, T., Gupta, K., Berger, I. & Rappsilber, J. Sequential digestion with Trypsin and Elastase in cross-linking/mass spectrometry. *Anal. Chem.* **7**, 4472-4478 (2019).
55. Giese, S. H., Fischer, L. & Rappsilber, J. A Study into the Collision-induced Dissociation (CID) Behavior of Cross-Linked Peptides. *Mol. Cell. Proteomics* **15**, 1094–1104 (2016).
56. Chen, F., Nielsen, S. & Zenobi, R. Understanding chemical reactivity for homo- and heterobifunctional protein cross-linking agents. *J. Mass Spectrom.* **48**, 807–812 (2013).
57. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Fragmentation behavior of a thiourea-based reagent for protein structure analysis by collision-induced dissociative chemical cross-linking. *J. Mass Spectrom.* **45**, 880–891 (2010).
58. Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).
59. Giese, S. H., Belsom, A., Sinn, L., Fischer, L. & Rappsilber, J. Noncovalently Associated Peptides Observed during Liquid Chromatography-Mass Spectrometry and Their Effect on Cross-Link Analyses. *Analytical Chemistry* **91**, 2678–2685 (2019).
60. Zhou, A. *et al.* Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *Elife* **4**, e10180 (2015).
61. Bertelsen, E. B., Chang, L., Gestwicki, J. E. & Zuiderweg, E. R. P. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8471–8476 (2009).
62. Choglay, A. A., Paul Chapple, J., Blatch, G. L. & Cheetham, M. E. Identification and characterization of a human mitochondrial homologue of the bacterial co-chaperone GrpE. *Gene* **267**, 125–134 (2001).
63. Braig, K., Adams, P. D. & Brünger, A. T. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nat. Struct. Biol.* **2**, 1083 (1995).
64. Wang, J. & Boisvert, D. C. Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)14 at 2.0A resolution. *J. Mol. Biol.* **327**, 843–855 (2003).
65. Clare, D. K. *et al.* ATP-Triggered Conformational Changes Delineate Substrate-Binding and -Folding Mechanics of the GroEL Chaperonin. *Cell* **149**, 113–123 (2012).
66. Nisemblat, S., Yaniv, O., Parnas, A., Frolow, F. & Azem, A. Crystal structure of the human mitochondrial chaperonin symmetrical football complex. *Proceedings of the National Academy of Sciences* **112**, 6044–6049 (2015).

67. Merle, N. *et al.* ATAD3B is a human embryonic stem cell specific mitochondrial protein, re-expressed in cancer cells, that functions as dominant negative for the ubiquitous ATAD3A. *Mitochondrion* **12**, 441–448 (2012).
68. Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F. U. Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**, 147–154 (1999).
69. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122**, 209–220 (2005).
70. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).